

# DM (and DAW) 2nd Session

Moderator(s): Suren Byna, Rob Ross

BOG Participants (one or both breakouts):

- Wes Bethel
- Rich Carlson
- Tina Declerck
- Adolfy Hoisie
- Latchesar Ionkov
- Bruce Jacob
- Terry Jones
- Scott Klasky
- Cara Leckey
- Jay Lofstead
- Carlos Maltzahn
- Ken Moreland
- Valerio Pascucci
- Tom Peterka
- Line Pouchard
- Lavanya Ramakrishnan
- Florin Rusu
- Christine Sweeney
- Matt Wolf
- Nick Wright
- John Wu
- Shinjae Yoo

# DM/DAW PRDs

1. **Workflow Execution.** EH systems and applications bring an unprecedented variety and number of data, resources, and services, bringing significant challenges to usability. **Research is needed to provide new methods of composing scientific workflows (e.g., via workflow motifs); to discover and map application workflow requirements to appropriate data, hardware, and software services; interfaces that facilitate HW and service composition; and to adapt computational strategy and data organization in response to resource availability, changing real-time constraints, and emerging results.** Note: the short title on this one needs a little more work.
2. **Programmability.** Software for (managing and understanding data in) EH systems, whether application software or system software, must be nimble with respect to the devices on which it can execute, the methods of communication, and robustness of the platform, and the relative performance of resources on which it is executing. **Research is needed into means for limiting the complexity of developing applications and services for these platforms, including approaches that enable better predictability (e.g., service level agreements); that allow trade-offs between compute, storage, and accuracy; that allow codes to better describe relationships between data and how they are to be used; that enhance scientific productivity; and that facilitate the validation of results.**
3. **EH Analytics.** (Data analytics) workflows on EH systems will require specialized algorithms and data models (for in situ analysis of multiple data modalities). **Research is needed to develop specialized, automated learning and analysis algorithms for use by domain scientists, not just data scientists; and to develop new algorithms and data models that exploit heterogeneous hardware (such as new storage and computing models) and to account for uncertainty in their methods and results.**
4. **Autonomy.** (Data management in service of) EH systems and workloads require complex balancing between a wide variety of storage and networking technologies in response to a widely varying workload and significant rate of faults. **Cross-cutting research is needed to understand what and how to instrument these systems; how to organize this telemetric data and make it available to multiple consumers; and how to analyze it in a manner that enables automated and goal-oriented decision-making (related to data location, organization, and transmission) under complex constraints and objectives.**